

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年12月19日(19.12.2024)



(10) 国際公開番号
WO 2024/257312 A1

(51) 国際特許分類:
G06N 3/08 (2023.01)

(21) 国際出願番号: PCT/JP2023/022294

(22) 国際出願日: 2023年6月15日(15.06.2023)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 富士通株式会社(FUJITSU LIMITED)
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小
田中4丁目1番1号 Kanagawa (JP).

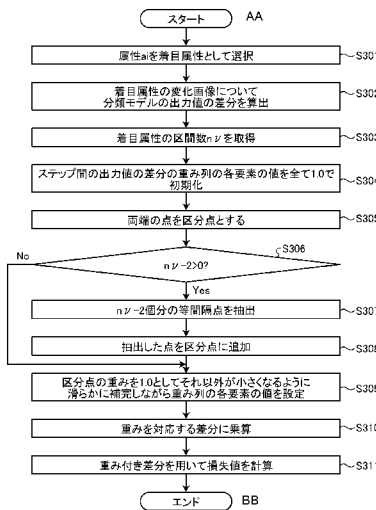
(72) 発明者: 吉井 章人(YOSHII, Akihito); 〒2118588
神奈川県川崎市中原区上小田中4丁目1番1
号 富士通株式会社内 Kanagawa (JP).

(74) 代理人: 弁理士法人酒井国際特許事務所
(SAKAI INTERNATIONAL PATENT OFFICE);
〒1000013 東京都千代田区霞が関3丁目8番1
号 虎ノ門ダイビルイースト Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC,
EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR,
HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,
KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,
LY, MA, MD, MG, MK, MN, MU, MW, MX, MY,
MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,

(54) Title: MACHINE LEARNING PROGRAM, MACHINE LEARNING METHOD, AND MACHINE LEARNING
DEVICE

(54) 発明の名称: 機械学習プログラム、機械学習方法及び機械学習装置



S301 Select attribute ai as attribute of interest
S302 Calculate difference in output value of classification model for changed image of attribute of interest
S303 Acquire number nv of intervals of attribute of interest
S304 Initialize, by 1.0, all of respective values of elements of weight column for difference in output value between steps
S305 Set both end points as division points
S307 Extract (nv - 2) equal interval points
S308 Add extracted points to division points
S309 Set respective values of elements of weight column while performing smooth complementation such that weight of division points is 1.0 and other weight is smaller than weight of division points
S310 Multiply corresponding difference by weight
S311 Use weighted differences to calculate loss value
AA Start
BB End

(57) Abstract: The present invention causes a computer to execute processes for: generating a plurality of pieces of changed data obtained by using a generation model to continuously change a specific attribute; determining whether the specific attribute is a discrete attribute or a continuous attribute; determining, when the specific attribute is determined to be a discrete attribute, error weights for a plurality of outputs, which are respectively obtained by inputting the plurality of pieces of changed data to a machine learning model, such that an error weight for a second output corresponding to no

SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

discrete boundary of the outputs is smaller than an error weight for a first output corresponding to the discrete boundary of the outputs; and training the machine learning model on the basis of a loss function including the determined error weights.

(57) 要約: 生成モデルを用いて特定の属性を連続的に変化させた複数の変化データを生成し、特定の属性が離散的な属性か連続的な属性かを判定し、特定の属性が離散的な属性と判定された場合、複数の変化データのそれぞれを機械学習モデルに入力して得られる複数の出力について、出力の離散的な境界に対応する第1の出力についての誤差の重みより、出力の離散的な境界以外に対応する第2の出力についての誤差の重みが小さくなるように決定して、決定した誤差の重みを含む損失関数に基づいて機械学習モデルを訓練する処理をコンピュータに実行させる。

明 細 書

発明の名称：

機械学習プログラム、機械学習方法及び機械学習装置

技術分野

[0001] 本発明は、機械学習プログラム、機械学習方法及び機械学習装置に関する。

背景技術

[0002] 近年、ディープラーニングによって訓練された機械学習モデルで意思決定支援をする試みが見られるようになってきている。このような意思決定支援において、バイアスと呼ばれる訓練データやモデル自身に含まれる偏りにより、機械学習モデルからの出力が公平な結果とならないことがある。公平性に影響を与える可能性が想定できる属性は、比較的容易に特定できる。ただし、このような公平性に影響を与える可能性が容易に想定できる属性以外にも公平性に影響を与える属性が解明できれば、不公平な結果をより適切に是正することができると考えられる。

[0003] そこで、従来、バイアスが存在する属性の特定の技術として以下の様な技術が提案されている。例えば、ある属性候補の軸に沿って変化を加えた変化画像の画像群を分類モデルで分類し、分類結果の変化する頻度と度合いを考慮して損失関数を生成する。そして、最適化により分類結果を変化させる傾向の高い、すなわちバイアスが存在する属性を得ることができる。

[0004] なお、機械学習の技術として、最小二乗確率的分類器により属性クラスの事後確率分布を求め、属性値を離散量として推定し、連続量スコア及び離散量スコアに対し重み付けをして統合する技術が提案されている。また、学習セットの各画像に付与された属性値を確率に変換し、評価値を画像毎に算出し、評価値が最小の推測モデルに対応する閾値に基づいて、対象事象の発現パラメータ値を決定する技術が提案されている。また、所定分布のサンプリングにより先行分類器のパラメータを初期化して偏りのない確率関数を用い

て学習を行い、先行分類器で学習されたパラメータにより後続分類器のパラメータを初期化して偏りのない確率関数を用いて画像を認識する技術が提案されている。また、バイアス感受性変数の確率分布に基づきバイアス除外モデルの訓練データを取得し、変数に基づき訓練データをクラスに分割して、変数の確率分布と一致するよう訓練データ集合の重みを調整する技術が提案されている。

先行技術文献

特許文献

[0005] 特許文献1：国際公開第2012/077476号

特許文献2：特開2020-135568号公報

特許文献3：米国特許出願公開第2019/0050689号明細書

特許文献4：米国特許出願公開第2020/0081865号明細書

非特許文献

[0006] 非特許文献1：Zhiheng Li, Chenliang Xu, Discover the Unknown Biased Attribute of an Image Classifier, In proceedings of ICCV (2021)

発明の概要

発明が解決しようとする課題

[0007] しかしながら、ある属性候補の軸に沿って変化を加えてバイアスが存在する属性を特定する技術では、属性が眼鏡をかけているか否か等のように離散的な意味を持つ場合に、余分に途中の変化を考慮してしまうため、属性発見の精度が低下する可能性がある。例えば、現実には考えられないような画像についても考慮してしまうおそれがある。そこで、離散的な意味を持つかどうかを手でラベル付けする方法も考えられるが、人手でのラベル付けは負担が大きく、データ数が多い場合には現実的ではない。さらに、連続的な意味を持ちうる属性でも2値ラベルが付加されている場合も存在するため、単に与えられたラベルからその属性が離散的であるか否かを区別することは困

難である。

[0008] また、最小二乗確率的分類器を用いて求めた属性クラスの事後確率分布から属性値を離散量として推定する技術では、属性クラスが最初から与えられており、クラスの数不明な属性が離散的か否かを判定することは困難である。また、学習セットの各画像に付与された属性値から得られる評価値を基に対象事象の発現パラメータ値を決定する技術では、属性が離散的か否かは考慮されていない。先行分類器で学習されたパラメータにより後続分類器のパラメータを初期化して偏りのない確率関数を用いて画像を認識する技術でも、属性が離散的か否かは考慮されていない。また、バイアス感受性変数の確率分布に基づき取得した訓練データを分割して訓練データ集合の重みを調整する技術でも、属性が離散的か否かは考慮されていない。したがって、いずれの技術を用いても、バイアスが存在する属性を特定することは難しく、機械学習モデルの推定精度を向上させることは困難であった。

[0009] 開示の技術は、上記に鑑みてなされたものであって、機械学習モデルの推定精度を向上させる機械学習プログラム、機械学習方法及び機械学習装置を提供することを目的とする。

課題を解決するための手段

[0010] 本願の開示する機械学習プログラム、機械学習方法及び機械学習装置は、一つの態様において、生成モデルを用いて特定の属性を連続的に変化させた複数の変化データを生成し、前記特定の属性が離散的な属性か連続的な属性かを判定し、前記特定の属性が離散的な属性と判定された場合、前記複数の変化データのそれぞれを機械学習モデルに入力して得られる複数の出力について、出力の離散的な境界に対応する第1の出力についての誤差の重みより、出力の離散的な境界以外に対応する第2の出力についての誤差の重みが小さくなるように決定して、前記決定した誤差の重みを含み損失関数に基づいて前記機械学習モデルを訓練する、処理をコンピュータに実行させる。

発明の効果

[0011] 本願の開示する機械学習プログラム、機械学習方法及び機械学習装置の一

つの態様によれば、機械学習モデルの推定精度を向上させることができるという効果を奏する。

図面の簡単な説明

- [0012] [図1]図 1 は、実施例に係る機械学習装置のブロック図である。
- [図2]図 2 は、出力値を示す数直線である出力値列の一例の図である。
- [図3]図 3 は、差分の重みの算出を説明するための図である。
- [図4]図 4 は、実施例 1 に係る機械学習装置による各属性の離散値を考慮した超平面生成処理の全体を示すフローチャートである。
- [図5]図 5 は、分類モデル訓練処理のフローチャートである。
- [図6]図 6 は、属性判定処理のフローチャートである。
- [図7]図 7 は、最適化処理のフローチャートである。
- [図8]図 8 は、実施例 2 に係る機械学習装置が属性判定処理で用いる座標平面を示す図である。
- [図9]図 9 は、機械学習装置のハードウェア構成図である。

発明を実施するための形態

- [0013] 以下に、本願の開示する機械学習プログラム、機械学習方法及び機械学習装置の実施例を図面に基づいて詳細に説明する。なお、以下の実施例により本願の開示する機械学習プログラム、機械学習方法及び機械学習装置が限定されるものではない。

実施例 1

- [0014] 図 1 は、実施例に係る機械学習装置のブロック図である。機械学習装置 1 は、制御部 10 及び記憶部 11 を有する。
- [0015] 記憶部 11 は、各属性について分類を行うための各属性それぞれの分類モデル 111 が集められた分類モデル群 110 を記憶する。分類モデル 111 は、機械学習モデルである。分類モデル群 110 は、分類モデル 111 のうちの 1 つとしてターゲット属性の分類を行うためのターゲット属性分類モデル 112 を含む。ターゲット属性とは、最終的に分類を行う対象とする属性であり、例えば、意思決定支援において判定に用いる対象となる属性である

。また、記憶部 11 は、属性の超平面に沿って連続した変化を示す変化画像を生成する画像生成モデル 120 を記憶する。

[0016] 制御部 10 は、ターゲット属性に対するバイアスとなる属性であるバイアス属性を特定して、バイアス属性の影響を軽減するようにターゲット属性の分類モデル 111 の訓練を行う。制御部 10 により訓練されたターゲット属性の分類モデル 111 を用いて分類を行うことで、バイアスを除去した公平な意思決定支援を行うことが可能となる。

[0017] 以下に、制御部 10 の詳細について説明する。制御部 10 は、分類モデル訓練処理部 101、属性判定部 102、最適化部 103 及び走査画像生成部 104 を有する。

[0018] 分類モデル訓練処理部 101 は、以下に説明する分類モデル訓練処理を実行して、分類モデル 111 のそれぞれを訓練する。分類モデル訓練処理部 101 は、訓練データセット 21 をデータセット格納装置 2 から取得する。また、分類モデル訓練処理部 101 は、訓練データセット 21 の訓練データが有する複数の属性の情報を含む属性集合を示すデータセット属性 22 を取得する。

[0019] 次に、分類モデル訓練処理部 101 は、データセット属性 22 で示される属性集合の中から着目属性を 1 つ選択する。次に、分類モデル訓練処理部 101 は、着目属性の分類モデル 111 を記憶部 11 が保持する分類モデル群 110 の中から特定する。そして、分類モデル訓練処理部 101 は、訓練データセット 21 に含まれる訓練データを用いて着目属性の分類モデル 111 の訓練を実行する。分類モデル訓練処理部 101 は、データセット属性 22 で示される全ての属性について分類モデル 111 の訓練を実行する。

[0020] 属性判定部 102 は、以下に説明する属性判定処理を実行して、各属性について離散的か連続的かを判定する。ここで、各属性に与えられるラベルの数を「分類数」と呼ぶ。例えば、メガネを掛けている顔かメガネを掛けている顔かという 2 つのラベルを有する属性は、ラベルは 2 値であり、分類数は 2 である。

- [0021] 本実施例では、離散的とする分類数の上限が予め決められており、属性判定部 102 は、離散的とする分類数の上限の情報を有する。例えば、分類数が 3 までが離散的とされた場合、分類数が 4 以上となる属性は連続的となる。
- [0022] 属性判定部 102 は、訓練データセット 21 及びデータセット属性 22 をデータセット格納装置 2 から取得する。そして、属性判定部 102 は、データセット属性 22 で示される属性集合の中から着目属性を 1 つ選択する。この場合、着目属性には、ターゲット属性は含まなくてよい。すなわち、着目属性は、ターゲット属性による分類においてバイアスとなる可能性のある属性である。
- [0023] 次に、属性判定部 102 は、訓練済みの着目属性の分類モデル 111 を記憶部 11 が保持する分類モデル群 110 の中から特定する。次に、属性判定部 102 は、訓練済みの着目属性の分類モデル 111 に訓練データを入力する。そして、属性判定部 102 は、着目属性の分類モデル 111 から出力された訓練データの個数分の出力値を判定用出力として収集する。
- [0024] ここで、属性判定部 102 は、出力値の取り得る範囲を 0～1 として正規化して、出力値を確率値に変換して判定用出力として属性判定処理に用いることもできる。出力値を正規化することで、異なる属性についても比較を容易に行うことが可能となる。
- [0025] 図 2 は、出力値を示す数直線である出力値列の一例の図である。図 2 における丸はそれぞれが各訓練データを入力した場合の出力値 130 を示す。図 2 に示すように、出力値 130 は、数直線上に表すことができる。図 2 は、出力値の取り得る範囲を 0～1 として、出力値を確率値として示した場合を示す。
- [0026] 次に、属性判定部 102 は、着目属性の分類数を n と設定して、以下の判定処理を $n = 2$ から 1 つずつ値を増やしていきながら実行する。属性判定部 102 は、収集した判定用出力について設定した分類数である n 個のグループへの分割を実行する。ここで、属性判定部 102 は、例えば、 k 近傍法等

のクラスタリング手法を用いて判定用出力のグループへの分割を行うことができる。

[0027] 次に、属性判定部 102 は、各グループの重心を算出する。属性判定部 102 は、重心として、各グループに含まれる判定用出力の平均や中央値等を用いることができる。そして、属性判定部 102 は、算出した各グループの重心がそれぞれ、出力値が取り得る値の範囲を n 等分した区間に入っているか否かを判定する。

[0028] ここで、特定の属性に基づく分類を行う分類モデル 111 の出力値が取り得る値の範囲は、画像分類であれば、特定の属性の超平面上での取り得る範囲内で値を変化させて変化画像を作成した場合の、両端の変化画像を入力とした場合の出力値の差分である。また、出力値を確率値とした場合であれば、出力値が取り得る値の範囲は $0 \sim 1$ となる。

[0029] 各グループの重心が n 等分した区間に入っている場合、属性判定部 102 は、着目属性は n 値的である離散的な属性と判定する。例えば、 n が 2 の場合、属性判定部 102 は、着目属性は 2 値的と判定する。

[0030] これに対して、各グループの重心が n 等分した区間に入っていない場合、属性判定部 102 は、 n を 1 つ増やして同様の判定処理を繰り返す。 n が分類数の上限に達しても各グループの重心が n 等分した区間に入っていない場合、属性判定部 102 は、着目属性を分類数の上限値以上の段階を有する連続的な属性であると判定する。

[0031] 例えば、図 2 において着目属性の段階が 4 の場合を考える。この場合、属性判定部 102 は、出力値 110 を 4 つのグループ 131 ~ 134 に分ける。そして、グループ 131 ~ 134 に含まれる判定用出力の重心を算出する。また、属性判定部 102 は、出力値が取り得る範囲である $0.0 \sim 1.0$ を 4 等分した区間 141 ~ 144 のそれぞれにグループ 131 ~ 134 の重心が含まれるか否かを判定する。この場合、属性判定部 102 は、区間 141 ~ 134 のそれぞれにグループ 131 ~ 134 の重心が含まれていないと判定する。

[0032] 属性判定部 102 は、ターゲット属性以外の属性の全てについて以上の判定処理を実行して離散的か連続的かを判定する。その後、属性判定部 102 は、各属性の判定結果を最適化部 103 へ出力する。この場合、属性判定部 102 は、離散的な属性については判定結果として分類数を通知する。また、属性判定部 102 は、連続的な属性については連続的であることを通知する。

[0033] 例えば、データセット属性 22 で示される属性集合に「年齢」、「性別」、「笑顔」及び「メガネ」の 4 つの属性が含まれる場合で具体例を説明する。ここでは、離散的とされる分類数の上限値が 3 である、すなわち分類数が 4 値以上である属性は、全て連続的であるとする。

[0034] 例えば、属性判定部 102 は、着目属性として「年齢」を選択する。次に、属性判定部 102 は、「年齢」に基づいて分類を行う分類モデル 111 に訓練データセット 21 に含まれる訓練データである画像を順次入力する。そして、属性判定部 102 は、各訓練データについて分類モデル 111 からの出力値を判定用出力として収集する。次に、属性判定部 102 は、判定用出力に対応する確率値を算出する。次に、属性判定部 102 は、分類数の初期値を 2 とする。そして、属性判定部 102 は、2 つのグループに判定用出力を分類する。そして、各グループの重心のそれぞれが属性に基づく分類の出力値の取り得る範囲を 2 等分割した区間のそれぞれに入っているか否かを判定する。ここでは、2 つのグループに属する判定用出力の重心が、2 つの区間のそれぞれに入らない場合を考える。この場合、属性判定部 102 は、その時点での分類数が 3 に達していないため、次の分類数として分類数を 3 とする。そして、属性判定部 102 は、3 つのグループに判定用出力を分類する。そして、各グループの重心のそれぞれが属性に基づく分類の出力値の取り得る範囲を 3 等分割した区間のそれぞれに入っているか否かを判定する。ここでは、3 つのグループに属する出力値の重心が、3 つの区間のそれぞれに入った場合を考える。この場合、属性判定部 102 は、着目属性である「年齢」の分類数が 3 値であり、連続的な属性であると判定する。

- [0035] このように、属性判定部 102 は、複数のデータを機械学習モデルである分類モデル 111 に入力して得られる判定用出力を前記特定の属性に基づいて所定数のグループに分割する。そして、属性判定部 102 は、グループ毎の判定用出力の重心が、特定の属性を変化させて得られる出力の取り得る範囲を所定数で等分割した分割範囲のそれぞれに含まれるか否を基に、特定の属性が離散的な属性か連続的な属性かを判定する。
- [0036] 最適化部 103 は、ターゲット属性以外の各属性の離散的か連続的な判定結果の入力を属性判定部 102 から受ける。そして、最適化部 103 は、離散的と判定された属性のそれぞれについて、差分に重みを付けて損失値を算出する以下に説明する損失値算出処理を実行して損失値を算出する。
- [0037] 最適化部 103 は、離散的と判定された属性の中から着目属性を 1 つ選択する。次に、最適化部 103 は、着目属性の変化データについて分類モデル 111 の出力値の差分を算出する。変化データは、属性が変化し得る範囲の中で、両端を基準として所定の数の変化の段階となるように属性を変化させた場合のデータである。すなわち、属性が変化しうる範囲は、隣り合う変化データ間の差分の総和にあたる。
- [0038] 例えば、画像認識による分類であれば、変化データは対象の属性を変化させた変化画像にあたる。変化データについての、変化の段階の数は予め決められている。以下では、変化データの数を「ステップ数」と呼ぶ。例えば、画像認識による分類において、メガネを掛けている状態とメガネをかけていない状態について考える。この場合、メガネをかけている画像及びメガネをかけていない画像が、属性が変化しうる範囲の中の両端の変化画像である。そして、ステップ数が 7 であれば、メガネをかけている画像及びメガネをかけていない画像を含む 7 枚の変化画像が存在する。
- [0039] 本実施例では、ステップ数は 4 より大きく設定される。最適化部 103 は、予め生成された変化画像を有しても良いし、画像生成モデル 120 を用いてその着目属性の超平面に沿って決められたステップ数に応じて値を変化させたデータを入力した場合に出力される変化画像を取得してもよい。

[0040] 図3は、差分の重みの算出を説明するための図である。ここでは、画像の分類を例に説明する。出力値列200は、1つ1つの変化画像を変化の順に入力した場合の出力値を順に並べた列を表す。この場合、ステップ数は7であり、7枚の変化画像を入力した場合の分類モデル111からの出力値201～207が存在する。出力値列200における各変化画像の出力値が位置する領域を「ステップ」と呼ぶ。出力値201～207は、7枚の変化画像に対応する。出力値201は最初の変化画像に対応する出力であり、出力値202～206は着目属性が徐々に変化する変化画像に対応する出力にあたり、出力値207は変化し終わった最後の変化画像に対応する出力である。

[0041] 次に、最適化部103は、 n 値のラベルを有する着目属性の分類数である n を、変化し得る範囲の区間数 n_{ν} とする。ここでは、着目属性の分類数 n が3である場合を例に説明する。すなわち、最適化部103は、区間数 n_{ν} を3とする。次に、最適化部103は、隣り合う変化データの間の差分の重み列の要素の値を全て1.0で初期化して、図3における状態211とする。図3の場合、ステップ数は7であるので、重み列の要素であるステップ間の差分は6個存在する。ここで、隣り合う変化データの間の差分とは取り合うステップにおける変化画像間の差分である。紙面に向かって出力値列200の下方に記載した値が、重み列の各要素の値にあたる。以下では、各ステップの間を「境界点」と呼ぶ。すなわち、境界点毎に重み列の要素が存在し、出力値列200における重み列の要素の位置を境界点とすることができる。

[0042] 次に、最適化部103は、出力値列200における境界点の両端の位置を区分点 $\{x\}$ として取得する。ここで、区分点 $\{x\}$ は、出力値列200を連続画像のステップではなく、本来の離散的な分類数により区分けした場合の境界にあたる。分類数が n であれば、区分点 $\{x\}$ は $n-1$ 個存在する。例えば、図3の出力値列200であれば、最適化部103は、状態212で示すように、重み列の要素221及び222に対応する境界点を区分点 $\{x\}$ に含ませる。この時、 n が2すなわち着目属性が2値的である場合には、

境界点は1つであるので区分点 $\{x\}$ も1つとなる。

[0043] 次に、最適化部103は、着目属性の段階が2であるか否か、すなわち区間数 n_v が2であるか否かを判定する。着目属性の段階が2である場合、最適化部103は、区分点 $\{x\}$ をステップ間の差分の重み列の両端の位置、実際には出力値列200の中央の点にあたる1つと決定する。

[0044] これに対して、着目属性の分類数 n が2より大きい場合、最適化部103は、区間数 n_v から境界点の両端の点を除いた $n_v - 2$ 個分の等間隔点を、出力値列200の中から抽出する。 $n_v = 3$ の場合、最適化部103は、1個の等間隔点を出力値列200の中から抽出する。

[0045] 図3では、ステップ数が7であるため等間隔点が境界点に一致しない。このように、等間隔点が境界点に一致しない場合が存在する。この場合には、 n が2の場合でステップ数が奇数の場合も含まれる。その場合、最適化部103は、等間隔点に境界点を区分点 $\{x\}$ として選択する。出力値列200において等間隔点が境界点のちょうど中間にあたる場合、最適化部103は、どちらの境界点を区分点 $\{x\}$ として選んでもよい。ここでは、最適化部103は、状態213に示すように、重み系列の要素223にあたる境界点を区分点 $\{x\}$ として選択する。

[0046] 区分点 $\{x\}$ の決定後、最適化部103は、決定した区分点 $\{x\}$ に位置する変化画像に対応する出力値の差分の重みを1.0として、それ以外が小さくなるように滑らかに補完しながら、状態214に示すように、重み系列の要素のそれぞれの値を決定する。ここでは、重み列の要素221、222及び223の値が1.0となるため、最適化部103は、それらの点を滑らかに結んで、重み列の要素224、225及び226の値を0.8と決定する。

[0047] この、区分点 $\{x\}$ が「出力の離散的な境界」の一例にあたる。そして、区分点 $\{x\}$ に位置する出力値の差分の重みが、「出力の離散的な境界に対応する第1の出力についての誤差の重み」の一例にあたる。また、それ以外の境界点に対応する出力値の差分の重みが、「出力の離散的な境界以外に対

応する第2の出力についての誤差の重み」の一例にあたる。すなわち、最適化部103は、特定の属性が離散的な属性と判定された場合、出力の離散的な境界に対応する第1の出力についての誤差の重みより、出力の離散的な境界以外に対応する第2の出力についての誤差の重みが小さくなるように決定する。より詳しくは、最適化部103は、変化させた順序にしたがい隣り合う変化データに対応する出力の差分それぞれを出力についての複数の誤差とする。そして、最適化部103は、変化データの順序及び出力の離散的な境界を基に、第1の出力についての誤差を出力についての複数の誤差の中から選択し、出力についての複数の誤差のうち第1の出力についての誤差以外を第2の出力についての誤差とする。

[0048] 次に、最適化部103は、変化画像に対応する出力値の差分の重みのそれぞれを対応する出力値の差分に乗算して重み付き差分を算出する。次に、最適化部103は、重み差分を用いて着目属性の損失値を計算する。例えば、非特許文献1の場合を想定すると、最適化部103は、Total variation及びOrthogonalization penaltyという2種類の損失値を算出する。Total variationは、着目属性を表す超平面での値を変化させて得られるTraversal画像を分類モデル111に入力して、分類結果の変化が大きいため公平性を満たしていないサンプルを求めるための損失値である。また、Orthogonalization penaltyは、潜在空間上のベクトルがターゲット属性と直交し、かつ既知の属性とも異なるように最適化するための損失値である。また、最適化部103は、連続的な属性については重みを付けない差分を用いて損失値を計算する。このように、最適化部103は、決定した誤差の重みを用いて損失関数に基づいて前記機械学習モデルを訓練する。

[0049] その後、最適化部103は、各属性についての算出した損失値を用いて損失関数になるべく小さくするように超平面を表すパラメータを調整して、各属性の超平面を生成する。以下では、パラメータを調整した超平面を、最適化された超平面と呼ぶ。その後、最適化部103は、各属性の最適化された超平面を走査画像生成部104へ出力する。

[0050] ここで、連続画像に対応する出力値の差分の重み付けについて、重み付けの付け方次第では、総和が重み付けしない場合よりも小さくなることが考えられる。この場合、同じ属性同士の比較であれば、重みが共通であるため問題とならない。これに対して、異なる属性同士の比較では、重み付けの仕方により、一方の属性の出力値が他方に比べて小さくなってしまふ等の意図しない結果となることが考えられる。

[0051] これらについては、重み付けを工夫することで解決することが可能である。1つの方法として、考慮するステップ数分の間の値を利用できるようにする方法がある。例えば、ステップ数が6、すなわち、境界点の数が5で3つの区分点{x}を考慮する場合、最適化部103は、区分点{x}に対応する重みとして、各境界点に対応する重みを1とした場合の合計である5を、区分点{x}の数である3で除算した $5/3$ を区間点{x}の重みとして割り当ててもよい。これにより、全体の重みの総和を一定にすることができる。この場合、ステップ数が3で連続画像を生成する場合とは異なり、等間隔で生成した5つの連続画像で、分類モデル111のパラメータの細かい調整を行うことが可能となる。

[0052] 他にも、最適化部103は、正規化等により各属性が取り得る値の範囲を揃えることができる。例えば、最適化部103は、属性毎の変化データ間の差分の総和を一定範囲の値に変換し、異なる属性同士の比較を可能とすることができる。具体的には、最適化部103は、属性毎の取り得る値の範囲を0~1として、出力値を確率値で表してもよい。

[0053] 例えば、データセット属性22で示される属性集合に「年齢」、「性別」、「笑顔」及び「メガネ」の4つの属性が含まれる場合で具体例を説明する。ここでは、属性の超平面に沿って変化する連続画像のステップ数が6であるとする。また、ここでは、各属性を比較し易くするため、出力値が確率値に変換された場合で説明する。さらに、属性判定部102により「年齢」の分類数が3値と判定された場合で説明する。

[0054] 最適化部103は、着目属性として「年齢」を選択する。次に、最適化部

103は、「年齢」の超平面に沿って値を変化させて得られる6枚の連続画像の隣り合う同士の確率値の差分を算出する。次に、最適化部103は、「年齢」の分類数が3を区間数 n として取得する。次に、最適化部103は、出力値列における境界点に対応する重み列の5つの要素の値を全て1.0で初期化する。次に、最適化部103は、境界点の両端の点を区分点 $\{x\}$ として選択する。次に、最適化部103は、区間数 n から2を減算した1個の点による等間隔点を出力値列における境界点の中から選択する。そして、最適化部103は、選択した境界点を区分点 $\{x\}$ に加える。そして、最適化部103は、区分点 $\{x\}$ とした3つの境界点に対応する重み系列の要素の値を1.0とする。さらに、最適化部103は、区分点 $\{x\}$ とした境界点の間の重み系列の要素の値を1.0よりも小さくなるように滑らかに補間しながら設定する。そして、最適化部103は、決定した重みを、連続画像を入力とした場合の出力値の差分に乗算して、「年齢」に沿った連続画像の重み付け差分を算出する。その後、最適化部103は、算出した重み付け差分を用いて「年齢」に基づく分類を行う分類モデル111の損失値を計算する。

[0055] 走査画像生成部104は、各属性の最適化された超平面の入力を最適化部103から受ける。そして、走査画像生成部104は、最適化された各属性の超平面に沿って変化させたデータを入力して画像生成モデル120に入力して、各属性の超平面に沿った変化画像である走査画像を生成させる。この際、走査画像生成部104は、予め決められたベクトルの初期値を画像生成モデル120に用い、且つ、予め決められたバイアス属性の情報を基に画像生成モデル120の訓練を行った後に、訓練後の画像生成モデル120を用いて走査画像を生成してもよい。そして、画像生成モデル120から出力された走査画像は、ターゲット属性分類モデル112に入力され、その出力を用いてバイアスを減らすためのターゲット属性分類モデル112のパラメータの調整が行われる。

[0056] 図4は、実施例1に係る機械学習装置による各属性の離散値を考慮した超

平面生成処理の全体を示すフローチャートである。次に、図4を参照して、実施例1に係る機械学習装置による各属性の離散値を考慮した超平面生成処理の全体的な流れを説明する。

- [0057] 分類モデル訓練処理部101は、訓練データセット21及びデータセット属性22を用いて、記憶部11に格納された分類モデル群110に含まれる各属性の分類モデル111に対して分類モデル訓練処理を実行する（ステップS1）。
- [0058] 属性判定部102は、訓練データセット21及びデータセット属性22と各属性の分類モデル111を用いて、ターゲット属性以外の属性それぞれが離散的か連続的かを判定する属性判定処理を実行する（ステップS2）。
- [0059] 最適化部103は、離散的な属性について、連続画像の間の差分及びステップ数、並びに、分類数を用いたステップ間重みを算出及び算出したステップ間重みを用いた損失値の算出を行う最適化処理を実行する（ステップS3）。
- [0060] 最適化部103は、最適化処理により求められた損失値を用いて各属性の最適化された超平面を生成する（ステップS4）。
- [0061] 走査画像生成部104は、バイアスとなり得る属性であるバイアス属性の超平面を画像生成モデル120に用いて各バイアス属性の超平面に沿った走査画像を生成する（ステップS5）。この後、生成された走査画像を用いてターゲット属性分類モデル112を訓練することで、ターゲット属性分類モデル112による推定においてバイアス属性による偏りを軽減することができる。
- [0062] 図5は、分類モデル訓練処理のフローチャートである。図5に示したフローは、図4におけるステップS1で実行される処理の一例を示す。次に、図5を参照して、分類モデル訓練処理の流れを説明する。
- [0063] 分類モデル訓練処理部101は、訓練データセット21をデータセット格納装置2から取得する（ステップS101）。分類モデル訓練処理部101は、訓練データセット21に含まれる訓練データが有する属性を示すデータ

セット属性 22 も取得する。

[0064] 次に、分類モデル訓練処理部 101 は、属性を表す変数である i を 0 に設定する（ステップ S102）。ここでは、データセット属性 22 で示される属性が L 個存在する場合で説明する。

[0065] 次に、分類モデル訓練処理部 101 は、データセット属性 22 で示される属性集合の中から着目属性として属性 a_i を選択する（ステップ S103）。

[0066] 次に、分類モデル訓練処理部 101 は、属性 a_i を分類するための分類モデル 111 である分類モデル m_i を記憶部 11 が保持する分類モデル群 110の中から特定する（ステップ S104）。

[0067] そして、分類モデル訓練処理部 101 は、訓練データセット 21 に含まれる訓練データを用いて分類モデル m_i の訓練を実行する（ステップ S105）。

[0068] その後、分類モデル訓練処理部 101 は、 $i = L - 1$ か否かを判定する（ステップ S106）。 $i = L - 1$ でない場合（ステップ S106：否定）、分類モデル訓練処理部 101 は、 i を 1 つインクリメントする（ステップ S107）。その後、分類モデル訓練処理部 101 は、ステップ S103 へ戻る。

[0069] これに対して、 $i = L - 1$ の場合（ステップ S106：肯定）、分類モデル訓練処理部 101 は、分類モデル訓練処理を終了する。

[0070] 図 6 は、属性判定処理のフローチャートである。図 6 に示したフローは、図 4 におけるステップ S2 で実行される処理の一例を示す。次に、図 6 を参照して、属性判定処理の流を説明する。

[0071] 属性判定部 102 は、データセット属性 22 に含まれるバイアス属性の中から着目属性として属性 a_i を選択する（ステップ S201）。

[0072] 次に、属性判定部 102 は、属性 a_i の分類モデル 111 である分類モデル m_i を記憶部 11 が保持する分類モデル群 110の中から特定する。属性判定部 102 は、分類モデル m_i に訓練データセット 21 に含まれる訓練デ

ータをそれぞれ入力する（ステップS 2 0 2）。

[0073] 次に、属性判定部 1 0 2 は、分類モデル m_i から出力された訓練データの個数分の出力値を収集して判定用出力とする（ステップS 2 0 3）。

[0074] 次に、属性判定部 1 0 2 は、着目属性の分類数である n を 2 に設定する（ステップS 2 0 4）。ここでは、離散的と判定する分類数の上限が N である場合で説明する。

[0075] 次に、属性判定部 1 0 2 は、 k 近傍法等のクラスタリング手法等を用いて、判定用出力を n 個のグループへ分割する（ステップS 2 0 5）。

[0076] 次に、属性判定部 1 0 2 は、各グループの重心がそれぞれ、出力値が取得する値の範囲を n 等分した区間に入っているか否かを判定する（ステップS 2 0 6）。

[0077] 各グループの重心が n 等分した区間に入っている場合（ステップS 2 0 6：肯定）、属性判定部 1 0 2 は、属性 a_i の分類数は n であり、離散的である判定する（ステップS 2 0 7）。

[0078] これに対して、各グループの重心が n 等分した区間に入っていない場合（ステップS 2 0 6：否定）、属性判定部 1 0 2 は、 $n = N$ か否かを判定する（ステップS 2 0 8）。

[0079] $n = N$ でない場合（ステップS 2 0 8：否定）、属性判定部 1 0 2 は、 n を 1 つインクリメントする（ステップS 2 0 9）。その後、属性判定部 1 0 2 は、ステップS 2 0 5 へ戻る。

[0080] これに対して、 $n = N$ の場合（ステップS 2 0 8：肯定）、属性判定部 1 0 2 は、着目属性の分類数が N 以上であり、連続的であると判定する（ステップS 2 1 0）。

[0081] 属性判定部 1 0 2 は、着目属性とする属性 a_i を変更して、ターゲット属性以外の属性の全てについて以上のステップS 2 0 1 ~ 2 1 0 で示した判定処理を実行して離散的か連続的かを判定する。

[0082] 図 7 は、最適化処理のフローチャートである。図 7 に示したフローは、図 4 におけるステップS 3 で実行される処理の一例を示す。次に、図 7 を参照

して、属性判定処理の流を説明する。

[0083] 最適化部103は、離散的なバイアス属性の中から着目属性として属性 a_i を選択する（ステップS301）。

[0084] 次に、最適化部103は、属性 a_i の変化画像について属性 a_i の分類モデル111である分類モデル m_i の出力値の差分を算出する（ステップS302）。

[0085] 次に、最適化部103は、属性 a_i の分類数である n を、変化し得る範囲の区間数 n_v として取得する（ステップS303）。

[0086] 次に、最適化部103は、出力値列の境界点に対応する差分の重み列の要素の値を全て1.0で初期化する（ステップS304）。

[0087] 次に、最適化部103は、出力値列における境界点の両端の位置を区分点 $\{x\}$ として選択する（ステップS305）。

[0088] 次に、最適化部103は、 $n_v - 2$ が0より大きいかな否か、すなわち、属性 a_i の分類数が2であるかな否かを判定する（ステップS306）。 $n_v - 2 = 0$ の場合（ステップS306：否定）、最適化部103は、既に選択した境界点の両端の位置にあたる点が区分点 $\{x\}$ に含まれる点の全てである決定してステップS309へ進む。分類数が2の場合、境界点の両端の位置にあたる点は出力値列の中央の点である。

[0089] 一方、 $n_v - 2$ が0より大きい場合（ステップS306：肯定）、最適化部103は、 $n_v - 2$ 個分の等間隔点を、出力値列における境界点の中から抽出する（ステップS307）。

[0090] そして、最適化部103は、抽出した点を区分点 $\{x\}$ に追加する（ステップS308）。

[0091] 次に、最適化部103は、決定した区分点 $\{x\}$ に位置する境界点に対応する重みを1.0として、それ以外の境界点に対応する重みが1.0より小さくなるように滑らかに補間しながら、各重み系列の要素の値を設定する（ステップS309）。

[0092] 次に、最適化部103は、各ステップ間の重みのそれぞれを変化画像の出

力値の差分に乗算して重み付き差分を算出する（ステップS 3 1 0）。

[0093] そして、最適化部 1 0 3 は、重み差分を用いて着目属性の損失値を計算する（ステップS 3 1 1）。

[0094] 以上に説明したように、本実施例に係る機械学習装置は、連続的に変化する出力値が得られるターゲット属性以外の属性について、離散的か連続的かを判定する。そして、本実施例に係る機械学習装置は、離散的な属性についてその分類数に応じた重みを生成して、出力値の差分に重みを付けて損失値を計算する。これにより、自動的に人手によりラベル付けした場合と同等の評価尺度による分類を行うことができ、公平でない結果をもたらす属性の中で、意味のある属性を発見し易くなる。したがって、機械学習モデルによる推定におけるバイアスの影響を軽減することができ、公平性が担保された推定の精度を向上させることができる。

実施例 2

[0095] 次に、実施例 2 について説明する。本実施例に係る機械学習装置 1 も図 1 のブロック図で示される。モデルや損失関数の種類により、ラベルが 2 値であっても分類モデル 1 1 1 からの出力が 1 つの場合と 2 つの場合が考えられる。例えば、2 値交差エントロピー (binary_crossentropy) を用いた分類では出力は 1 つであるが、2 クラスのクラス交差エントロピー (categorical_crossentropy) を用いた分類では出力が 2 つである。出力が 1 つの場合であれば、出力値の分布は図 3 に示したように数直線上で考えることができるが、出力が 2 つの場合は、2 次元上の点として考えることも可能である。

[0096] そこで、本実施例に係る機械学習装置 1 は、2 値のラベルに対する出力が 2 つの場合に、出力値の分布を 2 次元上の点として考える。本実施例に係る属性判定部 1 0 2 は、段階が 2 の属性を離散的な属性とし、段階が 3 以上の属性を連続的な属性とする。段階が 2 の属性としては、例えば、笑顔でない顔と笑顔の顔とを示すクラスや、メガネをかけていない顔とメガネを掛けている顔とを示すクラス等がある。ここでは、2 つのクラスをクラス A とクラス B とする。出力値は、クラス A の確率値とクラス B の確率値とで表される

。

[0097] 図8は、実施例2に係る機械学習装置が属性判定処理で用いる座標平面を示す図である。属性判定部102は、クラスAの確率値を $P(A)$ とし、クラスBの確率値を $P(B)$ として、図8に示すような座標平面を設定する。この場合、出力値は $P(A) + P(B) = 1$ となる。なわち、出力値は、図8に示す直線L上にプロットされる。そして、直線Lを2等分する中心点Pに対して、矢印D1の方向に存在する点はクラスAと判定される出力値を表す点であり、矢印D2の方向に存在する点はクラスBと判定される出力値を表す点である。

[0098] 属性判定部102は、直線L上でクラスタリングを行い、所定の基準を設けて分割された領域内に重心が入っているか否かを判定することで、属性の段階が2であるか否かを判定することができる。ここで、属性判定部102は、例えば、所定の基準を用いた分割方法として、直線Lの中心点Pから矢印D1側の領域と矢印D2側の領域とで2分割する方法を用いることができる。以上の方法を用いることで、属性判定部102は、2つの出力を有する場合にも、属性のラベルが2値であり離散的であるか、ラベルが2値以上であり連続的であるかを判定することができる。

[0099] 以上に説明したように、本実施例に係る機械学習装置は、2値を表す出力値として2つの値が出力される分類モデルであっても、その分類モデルが分類の対象とする属性が離散的か連続的かを判定することができる。したがって、2値を表す出力値として2つの値が出力される分類モデルを使用する場合にも、機械学習モデルによる推定におけるバイアスの影響を軽減することができ、公平性が担保された推定の精度を向上させることができる。

[0100] (ハードウェア構成)

図9は、機械学習装置のハードウェア構成図である。次に、図9を参照して、機械学習装置1の各機能を実現するためのハードウェア構成の一例について説明する。

[0101] 図9に示すように、機械学習装置1は、例えば、CPU (Central Proces

sing Unit) 91、メモリ92、ハードディスク93、ネットワークインタフェース94を有する。CPU91は、バスを介して、メモリ92、ハードディスク93及びネットワークインタフェース94に接続される。

[0102] ネットワークインタフェース94は、機械学習装置1と外部装置との間の通信のためのインタフェースである。ネットワークインタフェース94は、例えば、データセット格納装置2とCPU91との間の通信を中継する。

[0103] ハードディスク93は、補助記憶装置である。ハードディスク93は、図1に例示した記憶部11の機能を実現する。また、ハードディスク93は、図1に例示した、分類モデル訓練処理部101、属性判定部102、最適化部103及び走査画像生成部104の機能を実現するためのプログラムを含む核種プログラムを格納する。

[0104] メモリ92は、主記憶装置である。メモリ92は、例えば、DRAM (Dynamic Random Access Memory) を用いることができる。

[0105] CPU91は、ハードディスク93から各種プログラムを読み出して、メモリ92に展開して実行する。これにより、CPU91は、図1に例示した、分類モデル訓練処理部101、属性判定部102、最適化部103及び走査画像生成部104の機能を実現する。

符号の説明

- [0106]
- 1 機械学習装置
 - 2 データセット格納装置
 - 10 制御部
 - 11 記憶部
 - 101 分類モデル訓練処理部
 - 102 属性判定部
 - 103 最適化部
 - 104 走査画像生成部

請求の範囲

- [請求項1] 生成モデルを用いて特定の属性を連続的に変化させた複数の変化データを生成し、
- 前記特定の属性が離散的な属性か連続的な属性かを判定し、
- 前記特定の属性が離散的な属性と判定された場合、前記複数の変化データのそれぞれを機械学習モデルに入力して得られる複数の出力について、出力の離散的な境界に対応する第1の出力についての誤差の重みより、出力の離散的な境界以外に対応する第2の出力についての誤差の重みが小さくなるように決定して、前記決定した誤差の重みを含み損失関数に基づいて前記機械学習モデルを訓練する、
- 処理をコンピュータに実行させることを特徴とする機械学習プログラム。
- [請求項2] 前記決定の処理は、
- 変化させた順序にしたがい隣り合う前記変化データに対する出力の差分それぞれを出力についての複数の誤差とし、
- 前記変化データの順序及び出力の離散的な境界を基に、前記第1の出力についての誤差を前記出力についての複数の誤差の中から選択し、
- 、
- 前記出力についての複数の誤差のうち前記第1の出力についての誤差以外を前記第2の出力についての誤差とする、
- 処理を含むことを特徴とする請求項1に記載の機械学習プログラム。
- [請求項3] 前記判定の処理は、
- 複数のデータを前記機械学習モデルに入力して得られる判定用出力を前記特定の属性に基づいて所定数のグループに分割し、グループ毎の判定用出力の重心が、前記特定の属性を変化させて得られる出力の取り得る範囲を前記所定数で等分割した分割範囲のそれぞれに含まれるか否を基に、前記特定の属性が離散的な属性か連続的な属性かを判

定する、

処理を含むことを特徴とする請求項 1 に記載の機械学習プログラム

。

[請求項4]

生成モデルを用いて特定の属性を連続的に変化させた複数の変化データを生成し、

前記特定の属性が離散的な属性か連続的な属性かを判定し、

前記特定の属性が離散的な属性と判定された場合、前記複数の変化データのそれぞれを機械学習モデルに入力して得られる複数の出力について、出力の離散的な境界に対応する第 1 の出力についての誤差の重みより、出力の離散的な境界以外に対応する第 2 の出力についての誤差の重みが小さくなるように決定して、前記決定した誤差の重みを含む損失関数に基づいて前記機械学習モデルを訓練する、

処理をコンピュータが実行することを特徴とする機械学習方法。

[請求項5]

複数の属性のそれぞれに基づいて分類を行う機械学習モデルを記憶する記憶部と、

生成モデルを用いて特定の属性を連続的に変化させた複数の変化データを生成し、

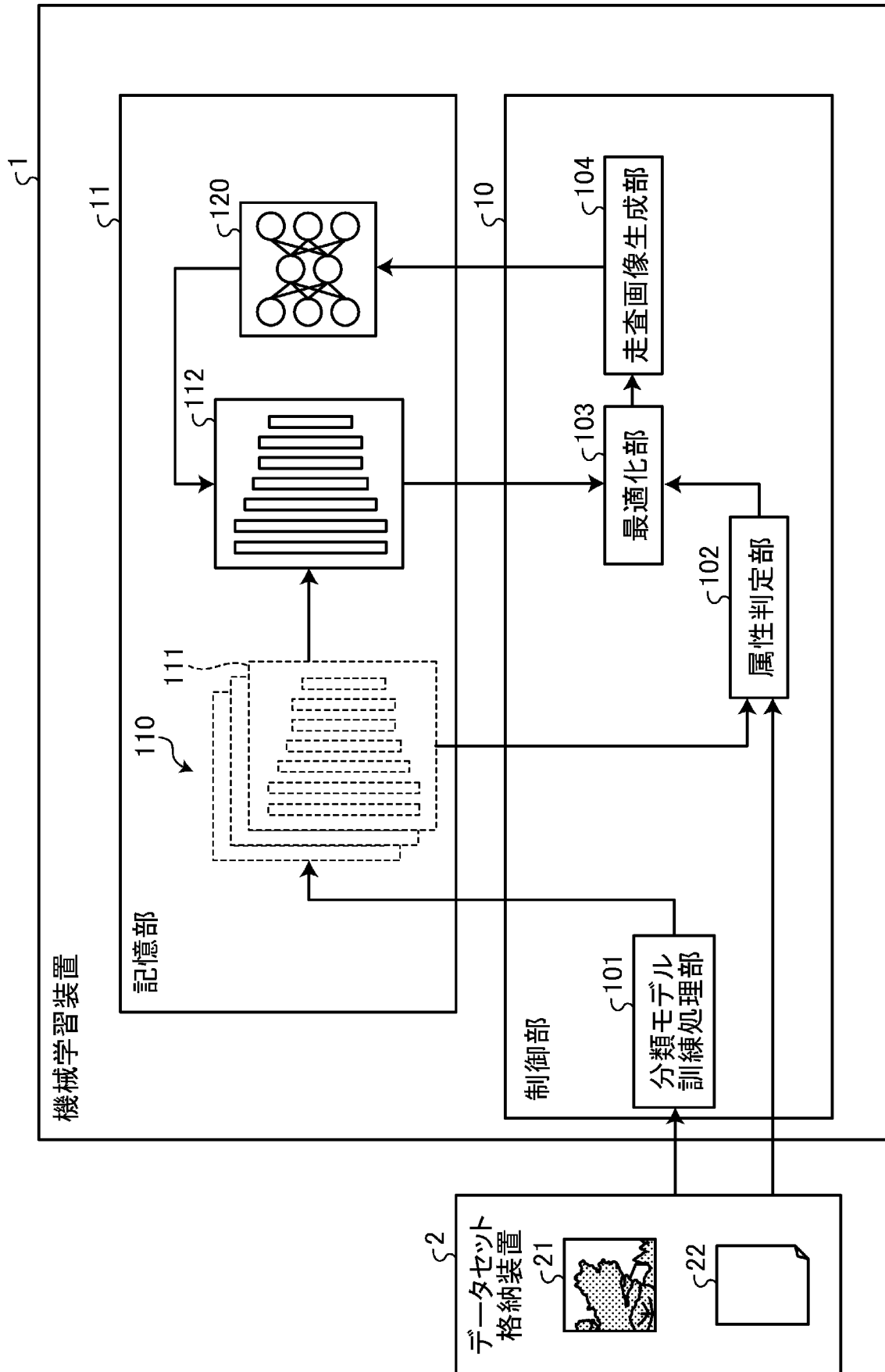
前記特定の属性が離散的な属性か連続的な属性かを判定し、

前記特定の属性が離散的な属性と判定された場合、前記複数の変化データのそれぞれを前記機械学習モデルに入力して得られる複数の出力について、出力の離散的な境界に対応する第 1 の出力についての誤差の重みより、出力の離散的な境界以外に対応する第 2 の出力についての誤差の重みが小さくなるように決定して、前記決定した誤差の重みを含む損失関数に基づいて前記機械学習モデルを訓練する、

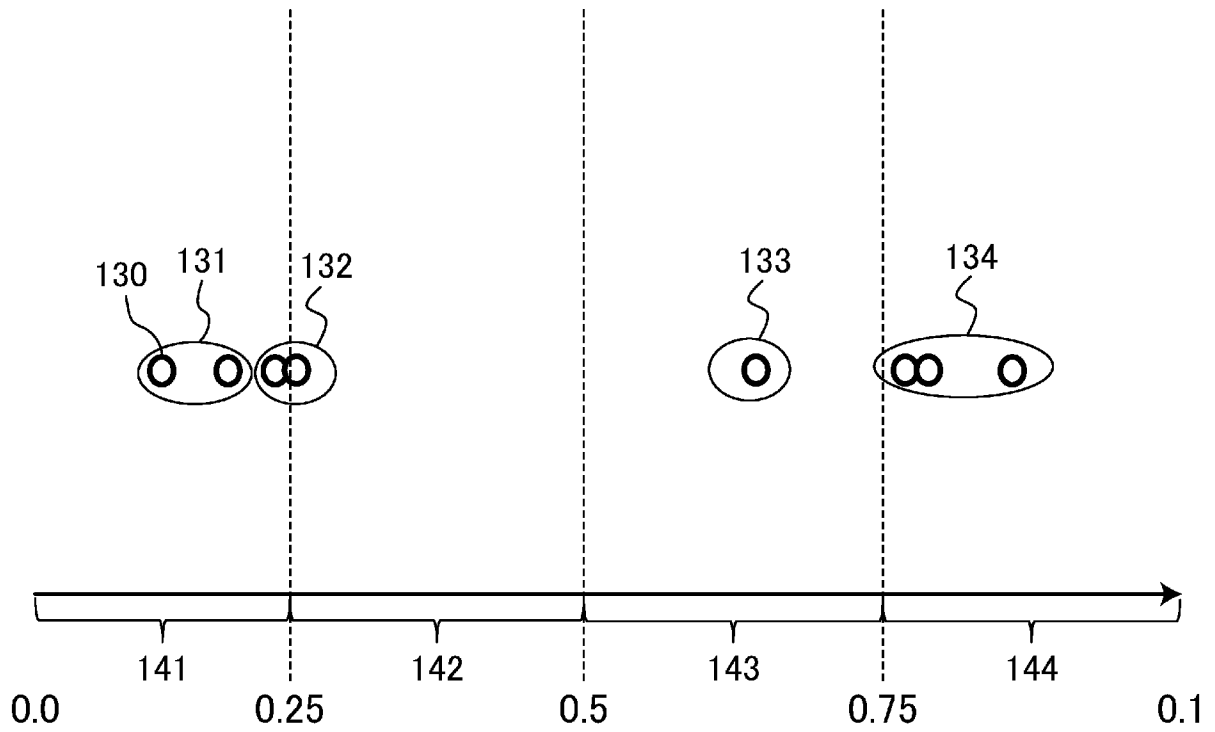
処理を実行する制御部と、

を備えたことを特徴とする機械学習装置。

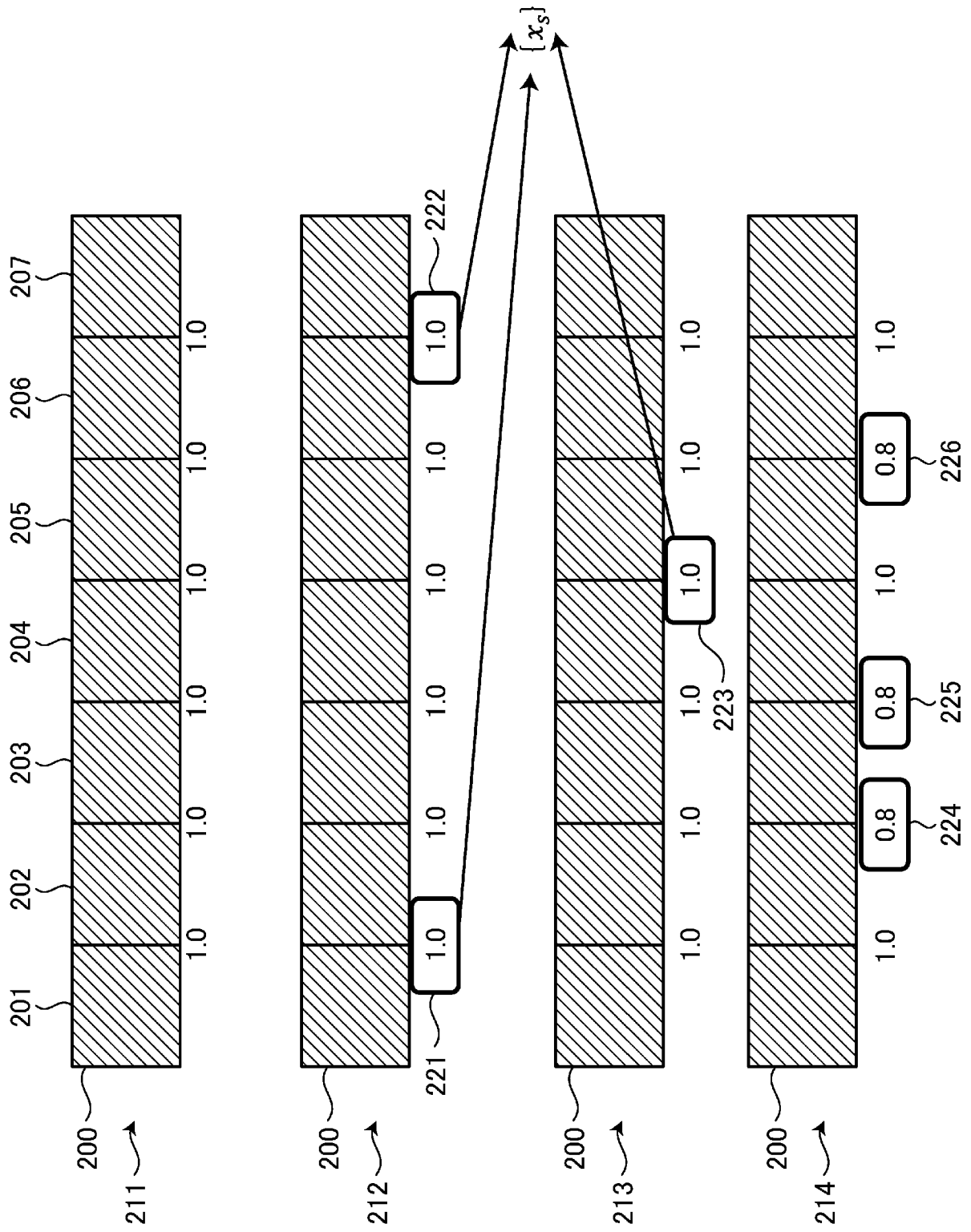
[図1]



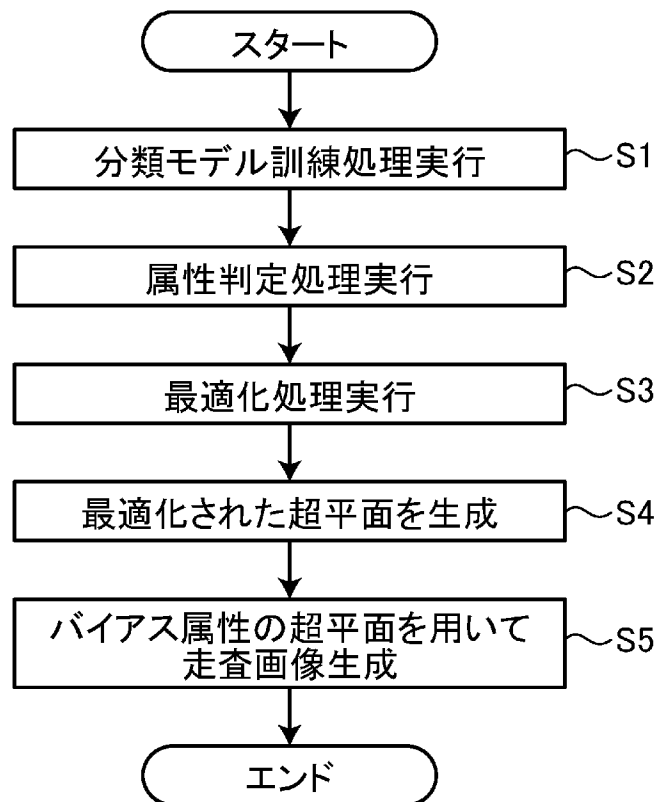
[図2]



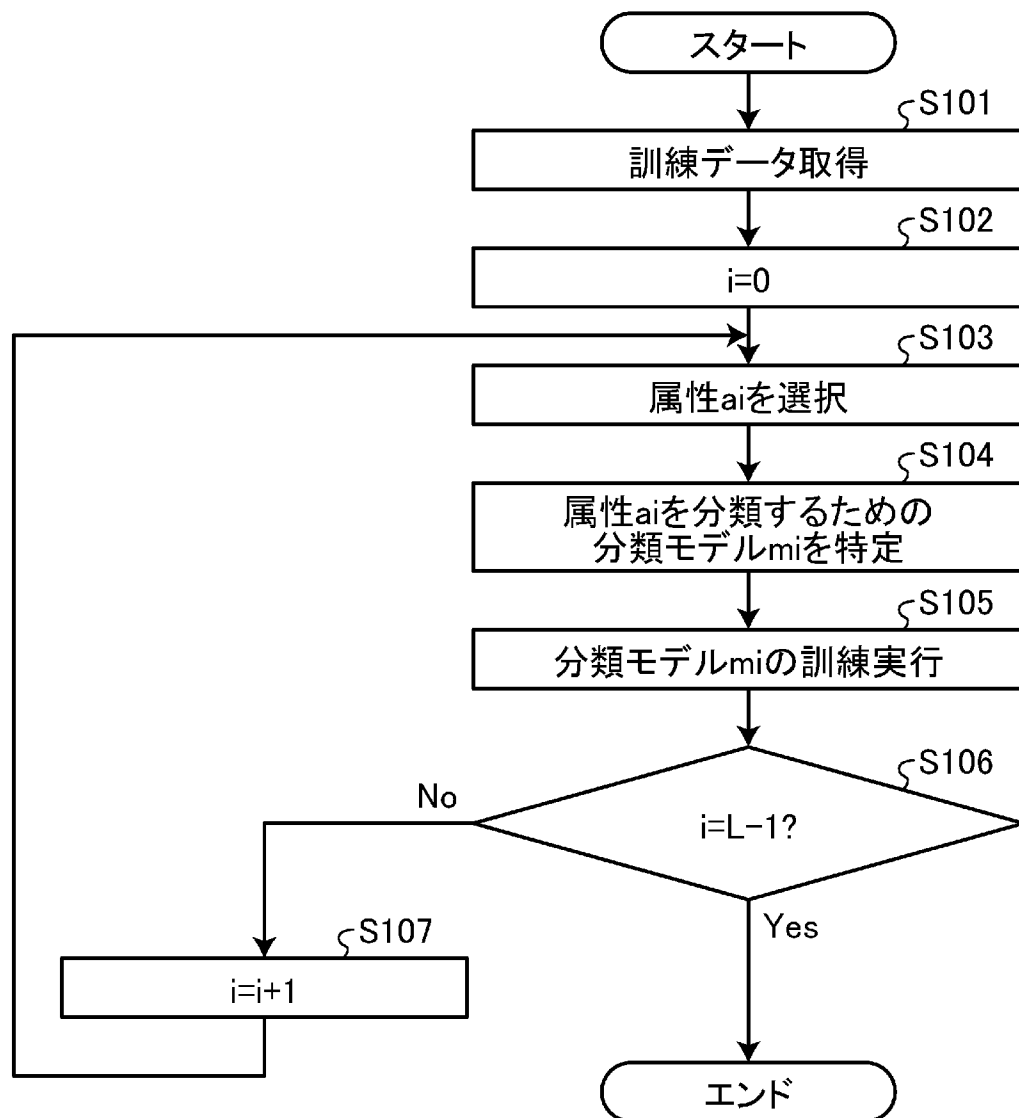
[図3]



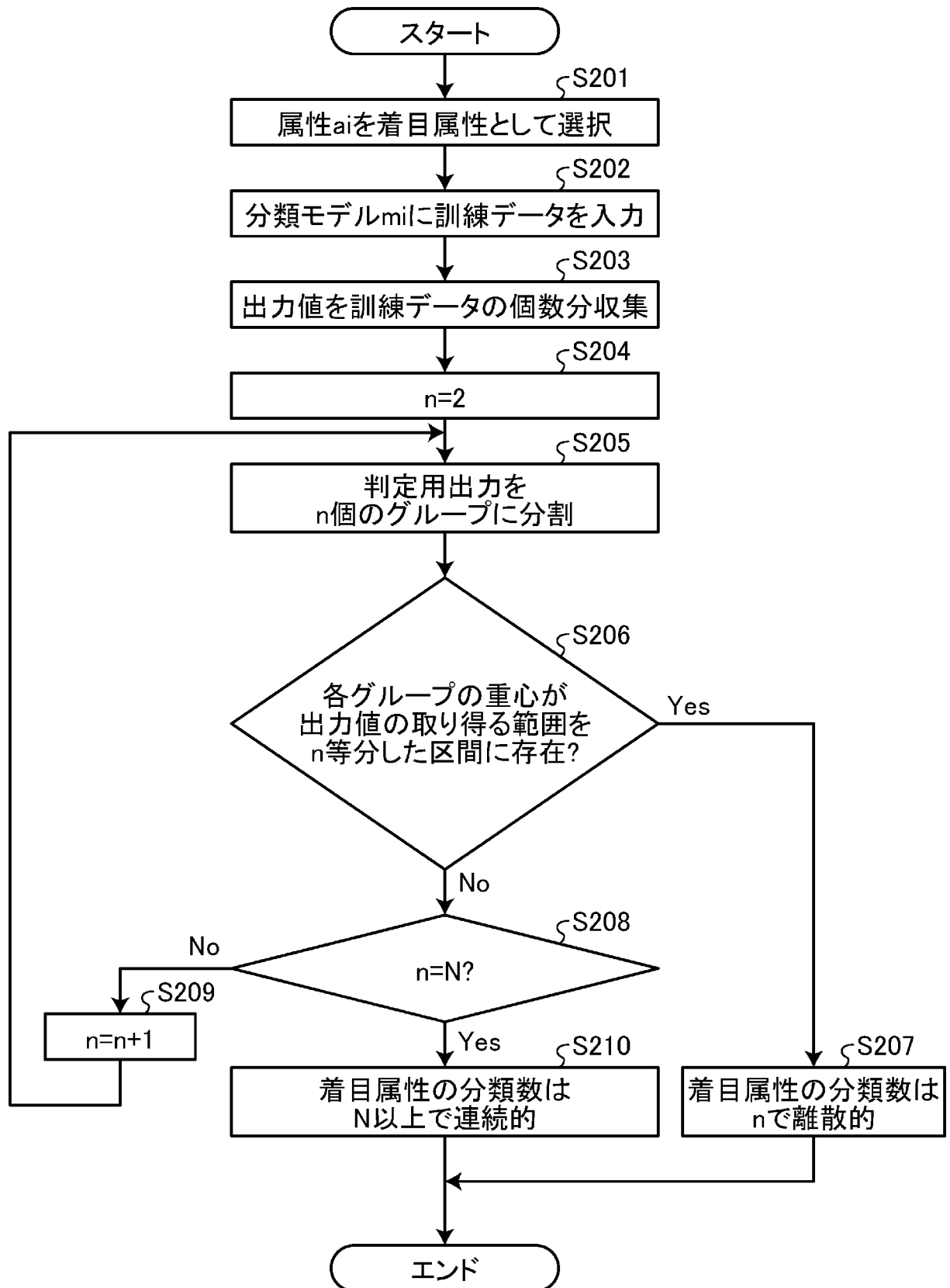
[図4]



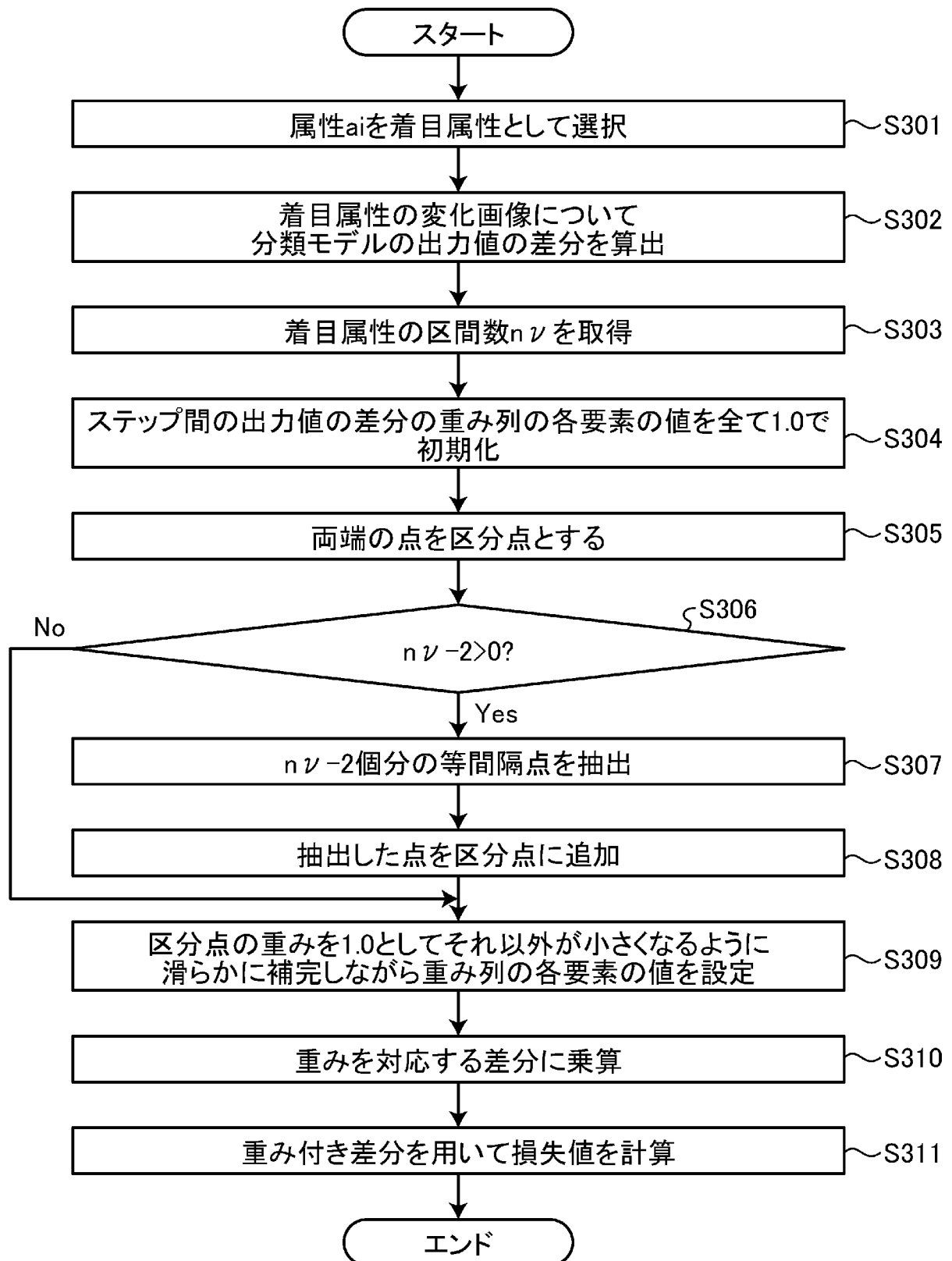
[図5]



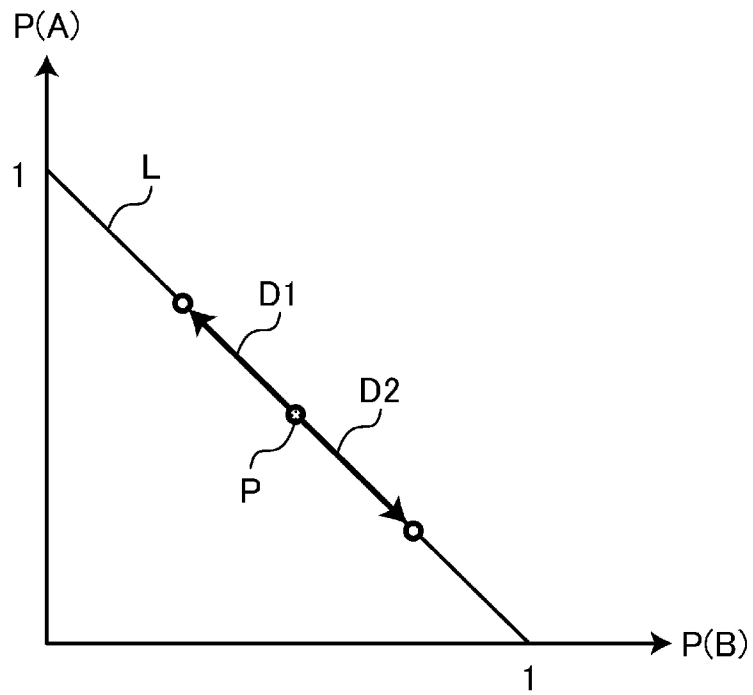
[図6]



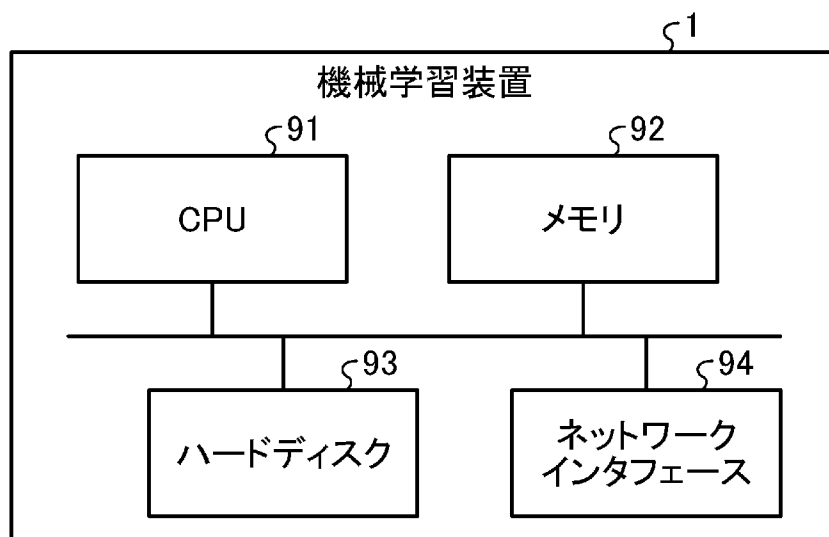
[図7]



[図8]



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/022294

A. CLASSIFICATION OF SUBJECT MATTER**G06N 3/08**(2023.01)i

FI: G06N3/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N3/00-20/00;

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2023

Registered utility model specifications of Japan 1996-2023

Published registered utility model applications of Japan 1994-2023

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HASSAN, S. et al., D-CALM: A Dynamic Clustering-based Active Learning Approach for Mitigating Bias, arXiv [online], 26 May 2023 [retrieved on 22 August 2023], all 12 pages, Retrieved from the Internet: <URL: https://arxiv.org/pdf/2305.17013.pdf > entire text	1-5
A	JP 2022-534160 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 28 July 2022 (2022-07-28) entire text	1-5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

22 August 2023

Date of mailing of the international search report

29 August 2023

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)**3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915****Japan**

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2023/022294

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2022-534160 A	28 July 2022	US 2021/0349920 A1 entire text	
		EP 3901789 A1	
		CN 110795638 A	
		KR 10-2021-0097204 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 3/08(2023.01)i FI: G06N3/08		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06N3/00-20/00; 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023年 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	HASSAN, Sabit et al., D-CALM: A Dynamic Clustering-based Active Learning Approach for Mitigating Bias, arXiv [online], 2023.05.26 [retrieved on 2023.08.22], all 12 pages, Retrieved from the Internet: <URL: https://arxiv.org/pdf/2305.17013.pdf> 全文	1-5
A	JP 2022-534160 A（ベイジン バイドウ ネットコム サイエンス テクノロジー カンパニー リミテッド）28.07.2022（2022-07-28） 全文	1-5
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献		
国際調査を完了した日 22.08.2023	国際調査報告の発送日 29.08.2023	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 今城 朋彬 5B 7888 電話番号 03-3581-1101 内線 3545	

国際出願番号
PCT/JP2023/022294

様式 PCT/ISA/210 (パテントファミリー用別紙) (2015年1月)